Scholar Publishing Group
International Journal of Neural Network
https://doi.org/10.38007/NN.2020.010204
ISSN 2790-2692 Vol. 1, Issue 2: 25-31

Analysis and Prediction of Tourism Data with Deep Learning

Xiahua Huang*

Yunnan Vocational and Technical College of Agricultural, Kunming, China

**corresponding author*

Keywords: Integrated In-depth Learning, Tourism data, Data Analysis, Prediction Research

Abstract: In recent years, deep learning (DL) technology has developed rapidly, which has been highly concerned by academia and industry. It has been widely used in image, voice, natural language processing, big data feature extraction and other aspects, and has also made very significant achievements. It is now a very important research direction in the field of artificial intelligence. The research on data analysis and prediction based on DL has also received more and more attention in recent years and has gradually become a research hotspot. Based on this, this fusion DL technology has studied the analysis and prediction of tourism data (TD). The training of DL prediction model network and TD are briefly analyzed; taking the TD of M province as the research object, this paper analyzes the number of tourists and hotels in the province and the experimental results verify the feasibility and effectiveness of applying the depth learning algorithm to the analysis and prediction of TD.

1. Introduction

With the continuous promotion of "Internet +", information technology has entered all walks of life, including tourism service industry. More and more people are used to obtaining more tourism service information from online platforms, so it is particularly important to help users quickly and accurately find out the tourism service project information they are interested in from the massive tourism service information. Therefore, this paper studies the analysis and prediction of TD by integrating DL.

Artificial neural network (NN) is very popular and widely used. As the basis of DL, its characteristics include the following: Strong learning ability. Like the brain, it can cope with the transformation of various inputs. When the input changes, it can adjust its extracted features in time, and has strong adaptability [1]. Parallelism; It can effectively detect the nonlinear relationship model of input to output, and externally, the NN is similar to a black box tool, hiding all other parts

of the NN structure except the input and output; Because there are many neurons in the NN, each neuron will share the contribution value, so each neuron has a relatively weak impact on the overall results [2].

This paper analyzes the tourist data, puts forward the DL technology, and studies the analysis and prediction of TD by integrating DL through the discussion of the algorithm and process of TD set fusion. Taking the TD of M province as an example, the paper conducts TD preprocessing and TD transformation, and finally verifies the feasibility and effectiveness of applying the depth learning algorithm to the analysis and prediction of TD through experimental tests [3-4].

2. Analysis of TD Integrated with DL

2.1. Analysis of Tourist Data

With the improvement of people's living standards, the demand for tourism is growing, enriching life, broadening vision and improving the quality of life. People go through a lot of thinking and design from deciding to travel to deciding which cities and scenic spots to visit, deciding when to travel, arranging itineraries, and arranging accommodation. During the period, the reference materials include oral recommendation from friends, browsing in the circle of friends, personal preferences, OTA evaluation, work arrangements, etc. When tourists determine the tourist cities and scenic spots, it is related to the source of this topic, that is, to choose the time of travel and then arrange accommodation [5-6]. When determining the travel time, passengers will consider the weather conditions of the day. Because the quality of weather will greatly affect the satisfaction of tourists. The next step is to arrange accommodation and choose a suitable hotel or guesthouse. There are many influencing factors. For this reason, this paper proposes to integrate DL technology to analyze and forecast TD [7].

2.2. Fusion Depth Model

2.2.1. Model Framework

As shown in Figure 1, it is the network framework of the DL prediction model proposed in this paper. The network model is composed of three parts, including user review IN, tourism service project review IN and other INs. These vectors can be regarded as various information expressions of users and tourism service projects. After the information is calculated through the network, each part will output a set of low dimensional dense space vectors, which can be regarded as further abstract description after feature extraction [8-9].

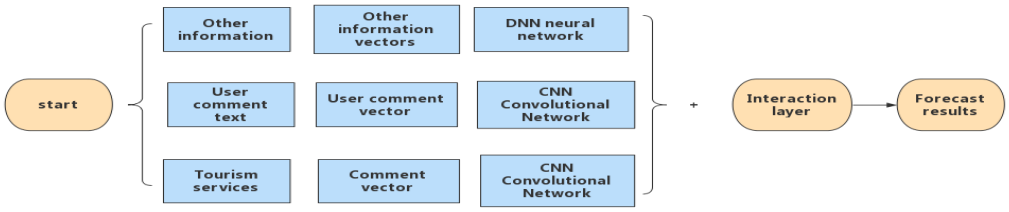

Figure 1. Framework diagram of depth prediction model

In the interaction layer, this paper uses the factor decomposition machine technology to build a model function to predict the score of the project. Because we need to excavate and learn the interactions between various vectors, predict and score the target project through these interactions, and this task can be well completed through the factor decomposition machine [10].

2.2.2. Training of DL Prediction Model Network

After defining the objective function, the next step is to optimize the weight value through the optimization algorithm. In the field of machine learning, stochastic gradient descent is a widely used weight optimization method, especially in DL. However, there are some problems with the random gradient descent method in practical applications [11-12]. Firstly, the random gradient learning descent method needs to choose a suitable learning rate; If the learning rate is too low, the update will be too slow and time-consuming; If the learning rate is too large, it may lead to failure of convergence, so it is difficult to choose an appropriate learning rate. In practical applications, for the loss function of non convex optimization (such as NN), it will converge to the local optimal solution in some cases, that is, it will be trapped at the saddle point. Therefore, in order to avoid these problems as much as possible, some algorithms based on random variable improvement and variation are usually used in practical applications [13].

3. Integrate DL Technology

3.1. DL Technology

NN is the basis of DL, which is also called artificial NN. Inspired by bionics research, it simulates the operation mode of brain neurons[14-15].The probability of the Skip gram model to predict the target vocabulary context vocabulary is calculated using the following formula (1):

$$p(context(k)\,|\,k) = \prod_{u \in contex(w)} p(s\,|\,k) \tag{1}$$

Contex (k) represents the context vocabulary of the target vocabulary k. Because the calculation efficiency is often very low when using Softmax to calculate p (s | k), Hierarchical Softmax is often used in existing models for efficient calculation[16].The document consists of n words in total, as shown in Formula (2).

$$T_{1:n}^{s} = \Phi(b_1^s) \oplus \Phi(b_2^s) \oplus \Phi(b_3^s) \oplus \Lambda \oplus \Phi(b_n^s) \tag{2}$$

For the comment information of tourism service projects, this method is also used[17].

3.2. Algorithm and Process of TD Set Fusion

Due to the large amount, variety and variety of TD, it is necessary to manually extract the required data set according to the needs of the subject before doing knowledge discovery, and then use the corresponding technical tools to integrate the data. After multiple database tables are obtained from TD through data analysis and statistics, the data set is fused into the data needed for the research according to the time dimension and hotel dimension association table. TDset fusion algorithm input: dataset D output: fused dataset MBEGIN.

The fusion algorithm of TD sets firstly classifies the data by subject manually, and each subject includes several data sets. The association and integration of data sets are performed on a per subject basis. Then further integrate the result set of each topic to get the final data set. There are two dimensions in the data, one is the hotel dimension and the other is the time dimension.

Fusion is divided into three steps. The first step is hotel dimension dataset fusion, the second step is time dimension dataset fusion, and the third step is time hotel dimension dataset fusion. Taking the smart TD of M Provincial Development Committee as an example, the steps are as follows.

The first step is the fusion of hotel dimension data sets. Correlate all hotel related data (except occupancy rate), including the name of the hotel, the number of scenic spots in a certain range

around the hotel, the number of hospitals in a certain range around the hotel, the number of convenience stores in a certain range around the hotel, and the number of sea areas in a certain range around the hotel. These data are fixed data and can be expanded, because the attributes that affect hotel occupancy are also in the tourism field. If there are still services for the hotel and facilities that interest tourists, you can continue to add them.

The second step is to use time as a foreign key, and use database technology to associate time related data in TD, including weather and wind, maximum temperature, minimum temperature, holidays, air quality index, and traffic flow. The associated time data table is shown in Table 1.

Table 1. Associated time data table

Attribute	Type	Length	Is It A Primary Key	Remarks
Time	Date	10	Yes	
Maximum Temperature	Int	2	No	Regression -- Wide Time Range
Minimum Temperature	Int	2	No	Regression -- Wide Time Range
Weather	Double(2,1)	2	No	Regression -- Wide Time Range
Wind Power	Double(2,1)	2	No	Regression -- Wide Time Range
Air Quality Index	Int	3	No	Fixed
Holiday And Vacations	Int	1	No	Regression -- 230 Items In July And August
Traffic Flow	Int	50	No	Remarks

The above data are all related to time, for example, the daily maximum temperature, daily minimum temperature, daily weather, daily wind, daily air quality index, and daily holiday code form a relationship model with time as the primary key.

4. Research on the Analysis and Prediction of TD based on DL

4.1. Analysis of TD in M Province

This paper takes the smart TD of Tourism Development Committee of M Province as an example to analyze TD. Data sources include public security, transportation bureau, OTA data, Internet, search engine, GIS, etc.TD cleaning. For example, the data of traffic flow on some dates are missing, which leads to inaccurate data generation results, and data cleaning is required. Take the smart TD of M Provincial Development Committee as an example to introduce the process of data clarity.

4.2. TD Transformation

4.2.1. Introduction to Data Transformation

TD transformation involves standardization and attribute construction. The weather conditions are quantified according to the impact of weather deterioration on tourists' tourism, as shown in Table 2.

Table 2. Weather data digitization

Serial no	Weather	Numerical
1	Sunny	1
2	Cloudy~sunny	1
3	Sunny~cloudy	1
4	Cloudy	1.5
5	Overcast~cloudy	1.5
6	Shower~clear	1.8
7	Shower~cloudy	2
8	Light rain~cloudy	2
9	Cloudy~thunder shower	2
10	Thunder shower~cloudy	2

The numerical weather conditions are defined according to the impact of actual weather deterioration on travel. For example, sunny days and cloudy sunny days, sunny cloudy days are the same for passengers, so they are all set as "1". The worse the weather is, the greater the value will be, which means the more negative the impact on the travel of passengers. Therefore, the passenger flow can be preliminarily predicted according to the weather conditions, thus affecting the hotel occupancy rate.

4.2.2. Holiday Digitization

Tourists will choose the right time to travel. Generally, tourists will choose holidays to travel, such as weekends, November, May Day, Tomb Sweeping Day, Mid Autumn Festival, New Year's Day and Spring Festival. Compared with students and teachers, summer vacation may be the preferred travel time. The numerical table of holiday data is shown in Table 3.

Table 3. Holiday data numerical table

Whether it is a holiday	Holiday Type	Numerical
Yes	Saturday and Sunday	1
No	Weekday	0
Yes	Legal holidays	2
No	Summer vacation	4
No	Winter vacation	4

Working days are working hours, and there will be fewer tourists, but it is not absolute, and some tourists will take holidays or ask for leave. Weekend is a day for people to travel in a short distance. Therefore, students and teachers may travel with their families. Therefore, businesses can design more packages on statutory holidays to satisfy passengers.

4.3. Experimental Results and Analysis

In order to illustrate the effectiveness of the recommended algorithm of the depth prediction model proposed in this paper, a comparative experiment with different depth model variants is designed. This paper uses MSE as the evaluation standard, and the results are shown in Figure 2:

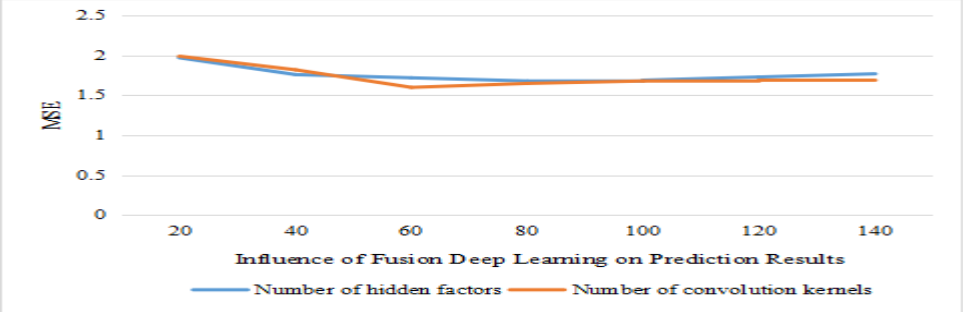

Figure 2. Influence of fusion DL on prediction results

From the results in Figure 2, we can see that when the number of convolution kernels reaches 100, the MSE of the model reaches the minimum. If it exceeds 100, the performance of the model will no longer improve. Therefore, it is more accurate to choose the tourism data analysis and prediction that integrates in-depth learning after comprehensive consideration.

5. Conclusion

This paper studies the analysis and prediction of TD by integrating DL technology, discusses the basic principle of NN, introduces common DL models, and verifies the effectiveness of this technology through experiments. However, this paper also has shortcomings. This paper does not take into account the impact of time factors when modeling users and tourism service information. Users will have different interests in different seasons. Time factors need to be taken into account when modeling, so that the model has timeliness, which can dig out more effective information and improve the accuracy of prediction; the social relationship between users is not considered, and the connection between users is ignored. According to the prediction results in this paper, the future research work will continue to study the combination prediction model, increase the accuracy of prediction, make reasonable predictions for various development trends, and put forward corresponding reasonable suggestions. Continue to deeply study the prediction of TD analysis, and find a more appropriate prediction model.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Yoo N, Lee E, Chung B J, et al. Analysis of Apartment Power Consumption and Forecast of Power Consumption Based on Deep Learning. Journal of IKEEE, 2019, 23(4):1373-1380.

[2] Brusaferri A, Matteucci M, Portolani P, et al. Bayesian deep learning based method for probabilistic forecast of day-ahead electricity prices. Applied Energy, 2019,

250(PT.1):1158-1175. https://doi.org/10.1016/j.apenergy.2019.05.068

[3] Safat W, Asghar S, Gillani S A. Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques. IEEE Access, 2020, PP(99):1-1. https://doi.org/10.1109/ACCESS.2020.3078117

[4] Ham Y G, Kim J H, Luo J J. Deep learning for multi-year ENSO forecasts. Nature, 2019, 573(7775):568-572. https://doi.org/10.1038/s41586-019-1559-7

[5] Benamrou B, Ouardouz M, Allaouzi I, et al. A Proposed Model to Forecast Hourly Global Solar Irradiation Based on Satellite Derived Data, Deep Learning and Machine Learning Approaches. Journal of Ecological Engineering, 2020, 21(4):26-38. https://doi.org/10.12911/22998993/119795

[6] Kovalev E V, Tverdokhlebova T I, Karpushenko G V, et al. Epidemiological situation of a new coronavirus infection (COVID-19) in the Rostov region: analysis and forecast. Medical Herald of the South of Russia, 2020, 11(3):69-78. https://doi.org/10.21886/2219-8075-2020-11-3-69-78

[7] Cubillos M. Multi-site household waste generation forecasting using a deep learning approach. Waste Management, 2020, 115(3):8-14. https://doi.org/10.1016/j.wasman.2020.06.046

[8] Bedi J, Toshniwal D. Deep learning framework to forecast electricity demand. Applied Energy, 2019, 238(MAR.15):1312-1326. https://doi.org/10.1016/j.apenergy.2019.01.113

[9] Mazaraki A, Boiko M, Okhrimenko A, et al. The impact of the national tourism system on the economic growth in Ukraine. Problems and Perspectives in Management, 2019, 17(4):93-101. https://doi.org/10.21511/ppm.17(4).2019.08

[10] Abdollahi M, Khaleghi T, Yang K. An integrated feature learning approach using deep learning for travel time prediction. Expert Systems with Application, 2020, 139(Jan.):112864.1-112864.11. https://doi.org/10.1016/j.eswa.2019.112864

[11] P Lara-Ben ítez, M Carranza-Garc á, J Garc á-Guti érrez, et al. Asynchronous dual-pipeline deep learning framework for online data stream classification. Integrated Computer Aided Engineering, 2020, 27(4):1-19. https://doi.org/10.3233/ICA-200617

[12] Gunter U, nder, Irem, Smeral E. Are Combined Tourism Forecasts Better at Minimizing Forecasting Errors? Forecasting, 2020, 2(3):211-229. https://doi.org/10.3390/forecast2030012

[13] Bhanja S, Das A. Deep Learning-based Integrated Stacked Model for the Stock Market Prediction. International Journal of Engineering and Advanced Technology, 2019, 9(1):5167-5174. https://doi.org/10.35940/ijeat.A1823.109119

[14] Pioppi B, Pigliautile I, Piselli C, et al. Cultural heritage microclimate change: Human-centric approach to experimentally investigate intra-urban overheating and numerically assess foreseen future scenarios impact. The Science of the Total Environment, 2020, 703(PT.2):134448.1-134448.15. https://doi.org/10.1016/j.scitotenv.2019.134448

[15] Mirzaev A. Perfection of the integral evaluation of the mechanism of recreational and tourist objects. Bulletin of Science and Practice, 2019, 5(2):127-134. https://doi.org/10.33619/2414-2948/39/17

[16] Tulsi P, Thakur D, Wen Y, et al. A Macro Analysis of Tourist Arrival in Nepal. Journal of Asian Finance Economics and Business, 2020, 8(1):207-215.

[17] Poghosyan K, Tovmasyan G. Modelling and Forecasting Domestic Tourism. Case Study from Armenia. SocioEconomic Challenges, 2020, 5(2):96-110. https://doi.org/10.21272/sec.5(2).96-110.2020

[18] Kropinova E G, Kuznetsova T Y, Fedorov G M. Regional Differences in the Level of Tourism Development in the Russian Federation. Geojournal of Tourism and Geosites, 2020, 32(4):1330-1336. https://doi.org/10.30892/gtg.32421-577